

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 07-281691

(43)Date of publication of application : 27.10.1995

(51)Int.Cl.

G10L 3/00

(21)Application number : 06-067251

(71)Applicant : NIPPON HOSO KYOKAI <NHK>

(22)Date of filing : 05.04.1994

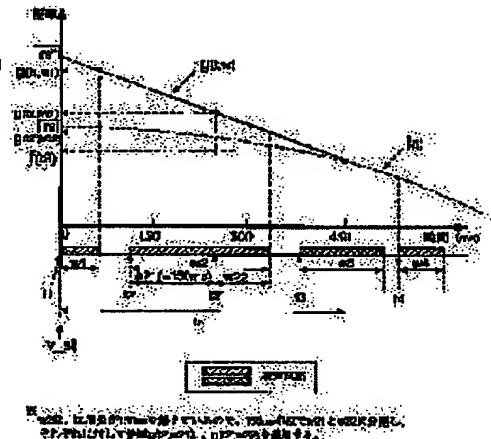
(72)Inventor : IMAI ATSUSHI
TSUGI TORU
NAKAMURA AKIRA
SEIYAMA NOBUMASA
MIYASAKA EIICHI

(54) SPEECH SPEED CONVERSION METHOD

(57)Abstract:

PURPOSE: To stably and naturally change the speech speed of inputted audio signals from 'slow' to 'fast'.

CONSTITUTION: While expanding the voiced sound segment of inputted audio signals in accordance with a certain rule, a expansion magnification function $g(t, w)$, that is a function of a segment length w of the voiced sound segment and time t at which the segment appears, is used to change the value of an expansion magnification. For example, the magnification function $g(t, w)$, which is shown in Figure 3 and is a function of w and t , is applied to a short voiced sound, which appears at a starting point of a phrase or appears ≤ 450 ms after a pitch change exceeds a certain value. If the length of a voiced sound segment is less than 150ms as w_1 shown in the figure, the magnification at the time of the end of a voiced sound segment is applied during the segment. On the other hand, if the voiced sound segment exceeds 150ms as w_2 , the segment is divided into 150ms units and the magnification which corresponds to the time of the respective ending point is applied.



(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平7-281691

(43) 公開日 平成7年(1995)10月27日

(51) Int.Cl.⁶

G 1 0 L 3/00

識別記号

H

庁内整理番号

F I

技術表示箇所

審査請求 未請求 請求項の数 7 O L (全 12 頁)

(21) 出願番号 特願平6-67251

(22) 出願日 平成6年(1994)4月5日

特許法第30条第1項適用申請有り 平成5年10月5日、
社団法人日本音響学会発行の「日本音響学会平成5年度
秋季研究発表会講演論文集-I」に発表

(71) 出願人 000004352

日本放送協会

東京都渋谷区神南2丁目2番1号

(72) 発明者 今井 篤

東京都世田谷区砧一丁目10番11号 日本放
送協会 放送技術研究所内

(72) 発明者 都木 徹

東京都世田谷区砧一丁目10番11号 日本放
送協会 放送技術研究所内

(72) 発明者 中村 章

東京都世田谷区砧一丁目10番11号 日本放
送協会 放送技術研究所内

(74) 代理人 弁理士 谷 義一 (外1名)

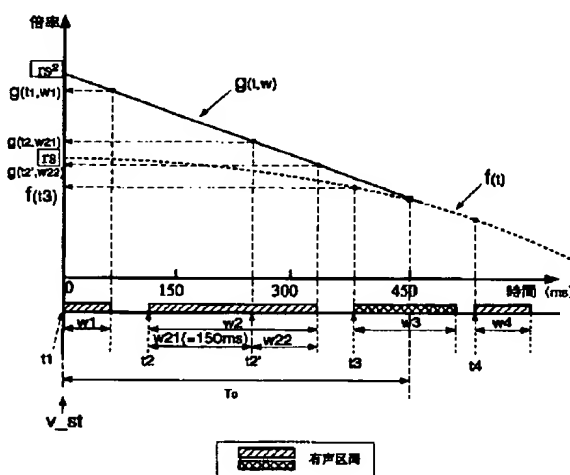
最終頁に続く

(54) 【発明の名称】 話速変換方法

(57) 【要約】

【目的】 入力音声の話速を「ゆっくり」から「速く」
に変化させる場合において、話速変化を安定、且つ自然
に実現すること。

【構成】 入力音声の有声区間を一定の規則に従って伸
張する際に、有声区間の区間長 w をその出現時刻 t の双
方を加味してその伸張倍率の値を変化させる倍率関数 g
(t, w)を適用する。例えば、フレーズの開始点やピ
ッチ変化がある一定の値を越えた時刻から450ms以
内に出現する150ms以下の短い有声音に対しては、
図3に示す倍率関数 $g(t, w)$ に沿って、有声音の長
さとその有声音の出現時刻の双方を加味した倍率を与
える。例として、図中の w_1 のように有声区間長が150
ms以下のものが存在する場合、有声区間の終了時刻に
おける倍率を、その有声区間に適用する。一方、 w_2
のように150msを越える長さの有声区間が存在する場
合には、150ms単位で分割し、各々の終了点に対応
する時刻での倍率を適用する。



■ w_2 は、区間長が150msを越えているので、150ms単位で w_2 と w_2 に分割し、
それぞれに対して倍率 $g(t_2, w_2)$ 、 $g(t_2, w_2)$ を適用する。

【特許請求の範囲】

【請求項1】 入力音声の無音区間、無声区間、有声区間を分離し、このうち有声区間を伸張することによって発声の速さ（話速）を声の高さを保ったまま遅くする変換を行う際に、各有声区間の時間長を逐次検出し、各々の有声区間の時間長に一樣な値の、あるいは経過時間とともに滑らかに変化する規準倍率を乗ずることにより、その倍率に対応した聴感的な効果を得る話速変換方法であって、一息で発声される区間（フレーズ）の一定時間内において、話速を「ゆっくり」の状態から「速い」状態に変化させる場合に、望まれる聴感上の話速変換効果において、1モーラ（1音）毎のテンポが自然に変化するように、各有声区間の時間長と、その有声区間の出現時刻の双方を考慮した一定の規則に基づいて、前記規準倍率に比べてより高い倍率を決定することを特徴とする話速変換方法。

【請求項2】 前記経過時間とともに滑らかに変化する基準倍率として、一息で発声する区間を単位にしてこの区間の開始点ではゆっくりとした話速を設定し、その終了点に向かって徐々に話速を速めることを特徴とする倍率関数を適用する場合に、上記区間の開始時刻から時間450ms以内に出現する150ms以下の短い有声区間に対しては、前記一定の規則に基づいて前記基準倍率に比べてより高い伸張倍率を供する倍率関数に沿って、その有声区間の時間長とその有声区間の出現時刻の双方を加味した伸張倍率を乗じ、また前記150msを越える時間長の有声区間に対しては、該150ms単位で分割し、各々の終了点に対応する時刻での伸張倍率を乗じ、さらに出現時刻が450msを越える有声区間に対しては前記規準倍率を乗ずることを特徴とする請求項1に記載の話速変換方法。

【請求項3】 前記一定の規則は、変換開始直後に出現する有声区間の伸張倍率が最も高く、時間経過と共に倍率は減少するが、相次いで現れる有声区間については、その時間長が短いものほど倍率の減少率は小さくするという規則であることを特徴とする請求項1に記載の話速変換方法。

【請求項4】 前記一定の規則は、任意の隣合うk番目、およびk+1番目の有声区間の時間長をそれぞれ w_k 、 w_{k+1} とし、話速変換後の時間長をそれぞれ w'_k 、 w'_{k+1} とすると、 w_k 、 w_{k+1} が近い値で且つ $w_k \leq w_{k+1}$ である場合のみ、 $w'_k > w'_{k+1}$ となることを許容することを特徴とする請求項1に記載の話速変換方法。

【請求項5】 前記一定時間は、ポーズ直後またはピッチの変化がある一定の値を越えた時刻から、フレーズの終了時刻までの間で任意に定める時間であり、所望の効果によってその時間の値を任意に決定できることを特徴とする請求項1に記載の話速変換方法。

【請求項6】 前記一定の規則は、当該有声区間の時間長が、ある一定の値よりも大きい場合には、所定の時間長単位で分割し、分割されたそれぞれの有声区間の時間長とその分割時刻の双方を考慮してそれぞれの倍率を決定するという規則であることを特徴とする請求項1に記載の話速変換方法。

【請求項7】 前記所定の時間長は、実用的な値の規準倍率によって有声区間を伸張したときに、聴感的な効果が感じとれなくなる最大の有声区間の時間長であることを特徴とする請求項6に記載の話速変換方法。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、話速変換方法に関し、特に聴覚障害者や高齢者等の音声補聴装置や、一般的な語学学習装置、ラジオ、テープレコーダー、電話などにおいて、話速変換による補助的聴取を行う際の聞き取り易さの向上、テレビジョン、ビデオテープレコーダー、ビデオディスクプレーヤーなどの音声出力を話速変換した際に生ずる映像と音声のズレを効果的に吸収するリアルタイム式の話速変換方法に関する。

【0002】

【従来の技術】従来の話速を変換する方式としては、有声区間を一樣倍率で伸張する手法と、有声区間を一樣倍率で伸張した場合に生ずる時間伸張を吸収する手法として、発声の開始点からの各有声区間の出現時刻のみの関数として倍率を単調に減少させるものがある。一樣伸張方式としては、ピッチ単位の波形繰り返しによるもの

（中村章ほか平成4年日本音響学会春季研究発表会「高品質リアルタイム話速変換システム」2-6-1P. 329-P. 330（1992-3））や、TDHS, P I C O L Aなどの手法を用いたものがある。時間伸張吸収方式は、音声の時間枠が限られている場合に、時間伸張を抑えつつ、一樣伸張による変換と同様の効果を実現しようとするもの（池沢龍ほか平成4年日本音響学会春季研究発表会「話速変換に伴う時間伸張を吸収するための一手法」2-6-2P. 331-P. 332（1992-3））である。これらの話速変換手法の基本的な考え方は、早口で発声された音声で、その区間長とは無関係に有声区間の一樣な倍率による伸張によって話速を「ゆっくり」にすることである。しかしながら、各々の有声区間長を考慮していないため、一樣な倍率を適用して全有声区間を伸張した場合においても、各有声区間長の差異によって聴感上の変換効果にばらつきを生ずることがあり、これが所望の話速変換効果に悪影響を及ぼすことがあった。ある一定の規則で話速を変化させる際に、有声区間長の差異に伴う聴感上の変換効果のばらつきを無くすために、各音韻長の時間的な大小関係を乱すことなく、所望の聴感的効果を実現する話速変換技術はなかった。

【0003】

【発明が解決しようとする課題】入力音声の話を「ゆっくり」にすることを目的として、無音区間、無声区間、有声区間を分離し、無音区間と無声区間の長さはそのままに、有声区間の伸張による話速変換を行う際、音声の中の複数の有声区間を一樣な倍率で伸張した場合、各々の有声区間の区間長によって、聴感上の「ゆっくり感」の程度に差異が生ずることがわかっている（今井篤ほか平成5年日本音響学会秋季研究発表会「話速変換に伴う時間伸張のリアルタイム吸収法」1-9-10P. 361-P. 362（1993-10））。

【0004】音声中には、異なる母音の連鎖や長母音などのように300msを越えるような比較的長い有声区間や、逆に無声区間や無音区間に挟まれた母音などに多い100msを下回るような比較的短い有声区間が相次いで現れることもあり、例えば、この両者が混在する音声に対して、一定の同じ倍率で伸張した音声を聴取した場合、長い有声区間は1有声区間単位での伸張時間の絶対量が大きく、聴感上の「ゆっくり感」が大きいのに比べ、短い有声区間は伸張時間の絶対量が小さく、場合によっては殆ど「ゆっくり感」が感じられないことがある。

【0005】例えば、区間長が350msと80msのものを従来法により一律に1.5倍に伸張した場合、525msと120msに変換されるが、前者の伸張時間の絶対増加量が175msであるのに対して、後者は僅か40msの伸張で、これが聴感上の効果の差となって現れてくる。従って、この様に長短さまざまな有声区間分布が一連の入力音声の中に複数箇所存在する場合は、話速の定まらない不安定な音声に変換されてしまい、場合によってはこれがかなり気になることがある。

【0006】また、既に提案されている、話速変換に伴う時間伸張を吸収する手法（池沢龍ほか平成4年日本音響学会春季研究発表会「話速変換に伴う時間伸張を吸収するための一手法」2-6-2P. 331-P. 332（1992-3））では、一息で発声する区間（フレーズ）の開始点での有声区間の伸張倍率を高く設定し、徐々に話速を速くしていくことで、変換音声の全体としての「ゆっくり感」と、全体としての時間伸張の吸収を実現しているが、このフレーズの開始点付近において短い有声区間が相次いで出現するような音声の場合には、上述の理由により比較的高い倍率を乗じても「ゆっくり感」が得られず、結果的に後半の話速の速い部分だけが目立ってしまうことになり、期待する効果が得られない場合がある。

【0007】上述した問題点を更に具体的事例で示す。

【0008】（1）一息で発声される区間（フレーズ）の予測長を2000msに固定し、伸張倍率 r を図1に示す曲線に添って rs （ $rs > 1$ ）から re （ $re < 1$ ）へと単調に減少させる。

【0009】（2）2000msを越えたところではピ

ッチ周波数の変化に伴い倍率に適宜修正を加える。

【0010】この手法をリアルタイム話速変換システムに導入し、多数のニュース音声を変換した結果、いくつかのフレーズについて期待される効果、特に、フレーズの開始点付近において「ゆっくり」した感覚を生じさせる効果の得られないものがあった。図2に、特に効果的であったフレーズ1例（同図の（a））と、特に効果が感じられなかったフレーズ2例（同図の（b）、

（c））について、フレーズ内の有声区間長の時間軸上の分布を示す。

【0011】この3例に代表される傾向として以下の点が挙げられる。

【0012】（1）フレーズの開始点付近450ms～500ms以内に150msを越える比較的長い有声区間が複数個存在する場合は、伸張倍率 r が $r = 1.4$ でも効果が大きい。

【0013】（2）フレーズの開始部分に150ms以下の比較的短い有声区間が存在する場合、 $r = 2.0$ でも効果が少ない。

【0014】他のフレーズについても検証した結果、同様の傾向が見られた。

【0015】また、一樣伸張の場合には、実用的な倍率を適用したときに、変換効果が感じられなくなるある一定の区間長以下の有声区間について、聴感上希望の変換効果が得られるように当該有声区間の区間長に応じて適用倍率を一時的に増大させることによって聴感上の変換効果のばらつきが効果的に解消できるが、話速を逐次変化させるような場合には、その時間経過に伴う倍率の変化と有声区間長の双方を考慮しなくてはならず、従来の話速変換方法では十分な変換効果が得られなかった。

【0016】本発明は、上述の問題点を鑑みてなされたもので、その目的は任意に定めた一定時間内において話速を「ゆっくり」の状態から「速い」状態に変化させる場合に、各音韻間の大小関係を保存し、各有声区間の変換音声の設定された話速に相応の聴感的効果を保ちつつ、安定した話速変換効果を得られるようにし、これにより聴覚障害者や高齢者に施す話速変換による補聴効果の向上や、定まった時間枠で出来るだけ自然なゆっくりとした音声を提供することが可能な話速変換方法を提供することにある。

【0017】

【課題を解決するための手段】上記目的を達成するために、本発明は、入力音声の無音区間、無声区間、有声区間を分離し、このうち有声区間を伸張することによって発声の速さ（話速）を声の高さを保ったまま遅くする変換を行う際に、各有声区間の時間長を逐次検出し、各々の有声区間の時間長に一樣な値の、あるいは経過時間とともに滑らかに変化する規準倍率を乗ずることにより、その倍率に対応した聴感的な効果を得る話速変換方法であって、一息で発声される区間（フレーズ）の一定時間

内において、話速を「ゆっくり」の状態から「速い」状態に変化させる場合に、望まれる聴感上の話速変換効果において、1モーラ（1音）毎のテンポが自然に変化するように、各有声区間の時間長と、その有声区間の出現時刻の双方を考慮した一定の規則に基づいて、前記標準倍率に比べてより高い倍率を決定することを特徴とする。

【0018】また、本発明を適用する前記経過時間とともに滑らかに変化する基準倍率として、一息で発声する区間を単位にして、この区間の開始点ではゆっくりとした話速を設定し、その終了点に向かって徐々に話速を速めることを特徴とする倍率関数を適用する場合に、上記区間の開始時刻から一定時間内、好ましくは時間450ms以内に出現する前記所定の長さに当る区間長、好ましくは150ms程度に満たない区間長を有する有声区間に対しては、前記一定の規則に基づいて前記基準倍率に比べてより高い伸張倍率を供する倍率関数に沿って、その有声区間の時間長とその有声区間の出現時刻の双方を加味した伸張倍率を乗じ、また前記150msを越える時間長の有声区間に対しては、該150ms単位で分割し、各々の終了点に対応する時刻での伸張倍率を乗じ、さらに出現時刻が450msを越える有声区間に対しては前記標準倍率を乗ずることを特徴とすることができる。

【0019】なお、上記の150ms、450msは好ましい値の一例を具体的に例示したもので、本発明はこれに限定されない。

【0020】また、本発明は好ましくは、前記一定の規則は、変換開始直後に出現する有声区間の伸張倍率が最も高く、時間経過と共に倍率は減少するが、相次いで現れる有声区間については、その時間長が短いものほど倍率の減少率は小さくするという規則であることを特徴とすることができる。

【0021】また、本発明は好ましくは、前記一定の規則は、任意の隣合うk番目、およびk+1番目の有声区間の時間長をそれぞれ w_k 、 w_{k+1} とし、話速変換後の時間長をそれぞれ w'_k 、 w'_{k+1} とすると、 w_k 、 w_{k+1} が近い値で且つ $w_k \leq w_{k+1}$ である場合のみ、 $w'_k > w'_{k+1}$ となることを許容することを特徴とすることができる。

【0022】また、本発明は好ましくは、前記一定時間は、ポーズ直後やまたはピッチの変化がある一定の値を越えた時刻から、フレーズの終了時刻までの間で任意に定める時間であり、所望の効果によってその時間の値を任意に決定できることを特徴とすることができる。

【0023】また、本発明は好ましくは、前記一定の規則は、当該有声区間の時間長が、ある一定の値よりも大きい場合には、所定の時間長単位で分割し、分割されたそれぞれの有声区間の時間長とその分割時刻の双方を考慮してそれぞれの倍率を決定するという規則であること

を特徴とすることができる。

【0024】また、本発明は好ましくは、前記所定の時間長は、実用的な値の標準倍率によって有声区間を伸張したときに、聴感的な効果が感じとれなくなる最大の有声区間の時間長であることを特徴とすることができる。

【0025】

【作用】本発明では、フレーズ内の一定の時間内に出現する有声区間の伸張倍率を、基本的にその出現時刻と区間長の双方を考慮した単調減少関数によって決定することにより、フレーズの開始直後に出現する有声区間の倍率が最も高く、時間経過と共に倍率は減少するが、相次いで現れる有声区間については、その区間長が短いものほど倍率の減少率は小さく、またある一定以上の長さを有する有声区間については予め定められた時間長単位で分割してそれぞれの区間長と分割時刻の両方を考慮して倍率を定める。これにより、本発明によれば、各フレーズ内の一定時間内において話速をゆっくりから速くに変化させる場合に、各音節間の音韻長の大小関係を乱すことなく、所望の変換効果を得ることが可能となる。この本発明の方法は従来の方法に加味することが可能である。例えば、有声区間を一樣伸張する際に、各フレーズの開始点や、ピッチの高さが一定値を越えたところから一定時間内において本発明を適用することは、フレーズの冒頭や、ピッチの変化の大きいところに相当する音節に対して安定した「ゆっくり感」を与えることが可能になり、高齢者や聴覚障害者に好適な「強調感」のある話速変換効果を得ることが可能となる。特に、従来提案されている話速変換による時間伸張を吸収する手法（池沢龍ほか平成4年日本音響学会秋季研究発表会「話速変換における時間伸張吸収のリアルタイム化の検討」2-9-2P. 349-P. 350（1993-10））に適用した場合は、発声の開始点付近の「ゆっくり感」が聴感上不安定であったという欠点が解消されることに加えて、比較的高い倍率から低い倍率に急激に変化させた場合においても、音韻長の時間的大小関係を乱すことなく所望の話速変換効果を供することが可能となる。更にこの「強調感」のため、従来方法に比べて話速変換倍率を低めに設定した場合でも、聴感上安定した話速変換効果が得られ、時間伸張率を抑えることも可能となる。

【0026】

【実施例】以下、図面を参照して本発明の実施例を詳細に説明する。

【0027】一息で発声されると予測される区間内で、この区間の開始点に於いては原音声の話速より「ゆっくり」とした話速を設定し、終了点に向かって一定の規則に従って話速を速めていく「リアルタイム話速変換方法」の手法（池沢龍ほか平成4年日本音響学会秋季研究発表会「話速変換における時間伸張吸収のリアルタイム化の検討」2-9-2P. 349-P. 350（1993-10））に本発明を適用した場合の実施例について

説明する。この「リアルタイム話速変換方法」は、実時間で動作する話速変換装置において、時間伸張を吸収する手法であるが、これは各フレーズの開始点においての聴感上違和感のない「ゆっくり感」を与えることがそれぞれのフレーズ全体での「ゆっくり感」の印象を決定するものであるため、これに本発明による方法を適用することは特に効果的であるといえる。

【0028】図3は本発明の一実施例の動作の概要を示す。フレーズの開始点付近から450ms以内に出現する150ms以下の短い有声音に対しては、図3に示す倍率関数 $g(t, w)$ に沿って、有声音の長さとその有声音の出現時刻の双方を加味した倍率を与える。一例として、図中の w_1 のように有声音区間長が150ms以下のものが存在する場合、有声音区間の終了時刻における倍率を、その有声音区間に適用する。一方、 w_2 のように150msを超える長さの有声音区間が存在する場合には、150ms単位で w_{21} と w_{22} に分割し、各々の終了点に対応する時刻での倍率を適用する。ただし、経過時間が450msを越えた時点で従来の伸張倍率曲線 $f(t)$ (図1)を適用する。

【0029】連続した短い有声音区間A、Bに対し、後続する有声音区間Bの継続時間 T_b が $T_b < T_a$ の場合に、Bの伸張倍率がAの倍率を上回ると、音韻長の知覚的バランスがくずれて聞きにくくなる場合がある。これに対し、本発明実施例ではこのような減少倍率の逆転現象は生じない。また、300msを越えるような長い有声音区間の場合、150ms単位で伸張倍率を減少させるため、いわゆる「まのび」した感覚を抑制する効果がある。

【0030】図4～図7は本発明の一実施例を更に詳細に示す図である。

【0031】図4は本発明の一実施例の全体回路構成を示すブロック図である。

【0032】図4に示すリアルタイム話速変換装置は、音声入力回路1と、CPU(中央処理ユニット)回路2と、PROM(プログラマブルROM)回路3と、入力バッファ回路4と、処理バッファ回路5と、ファイル回路6と、音声出力回路7と、バス8とを備えている。そして、音声入力回路1によって話速変換対象となる音声(原音声)を取り込み、リアルタイム処理で、原音声の

声の高さ(ピッチ周波数)の変化を検出すると共に、この検出結果に基づいて、声の高さの高い部分では話速を緩め、低い部分では話速を速めるという規則で話速を変化させることにより、原音声の発話時間を保ったまま、原音声を聴き易い良好な音声に変換する。

【0033】音声入力回路1は、原音声を入力するための一般的な構成の回路、例えばマイクロフォン、音調回路、A/D(アナログ/デジタル)変換器、音声記憶再生回路、音声記憶媒体(例えば、ICメモリ、ハードディスク、フロッピーディスクまたはVTR(ビデオテー

プレコーダ))、およびインターフェイス回路等を備えており、話速変換対象となる音声を取り込み、これをデジタル形式の音声信号に変換するとともに、この変換したデジタル音声信号をCPU回路2からの指示に基づいてフレーム単位で入力バッファ回路4に供給する。

【0034】入力バッファ回路4は、必要な容量のRAM(ランダムアクセスメモリ)などによって構成され、CPU回路2の作業域として使用される部分であり、音声入力回路1から出力される音声信号を取り込んでこれを記憶するとともに、CPU回路2からの指示に基づいて記憶している音声信号を処理バッファ回路5に転送する。

【0035】処理バッファ回路5は、必要な容量のRAMなどによって構成され、CPU回路2の作業域として使用される部分であり、入力バッファ回路4から出力される音声信号を取り込んでこれを記憶するとともに、CPU回路2からの指示に基づいて記憶している音声信号をファイル回路6などに転送する。

【0036】ファイル回路6は、RAMのほかに、ICメモリやフロッピーディスク等の音声記憶媒体によって構成され、本発明に係わる有声音区間の伸張された音声信号と、無音区間の短縮の処理を施された信号などを格納するメモリであり、処理バッファ回路5から処理済の音声信号が出力されたとき、これを取り込んで記憶し、この後CPU回路2からの指示に基づいて記憶している音声信号を音声出力回路7に供給する。

【0037】音声出力回路7は、ファイル回路6内の音声信号を外部に出力するための一般的な構成の回路、例えばインターフェイス回路、D/A(デジタル/アナログ)変換器、スピーカ、録音装置(あるいは放送機器)等を備えており、ファイル回路6から音声信号が出力されたとき、これを取り込んで音声に変換しながら、外部に出力する。

【0038】また、CPU回路2は、ワンチップマイクロコンピュータ等によって構成される部分であり、PROM回路3に格納されている図5、図6に示すようなプログラムに基づいて装置全体の制御や各種のデータ処理を行う。

【0039】また、PROM回路3は、CPU回路2の動作を規定するプログラムや各種の処理で使用される定数データなどの格納場所として使用される部分であり、CPU回路2からの読みだし指令に応じて記憶しているプログラムや定数データを読みだしてCPU回路2に供給する。

【0040】次に、本発明の一実施例の動作について図5、図6を参照して説明する。

【0041】図5、及び図6は処理の流れを示すフローチャートであり、図6は図5のST9の有声音区間処理ルーチンの詳細を示す。

【0042】ここでは、説明のために音声信号中の息継

ぎ区間を「ポーズ」、一息で発声される区間を「フレーズ」、また「フレーズ」の時間長の平均的な値を「予測フレーズ長」呼び、次のように定義する。

【0043】ポーズ：無音部分と判定された区間のうち、その区間長が T_{h1} （本実施例では $T_{h1} = 200 \text{ ms}$ ）以上の無音区間。なお、 T_h はスレッシュホールド値を意味する。

【0044】フレーズ：ポーズと次のポーズに挟まれる区間。

【0045】この区間の開始点を P_{h_st} とする。

【0046】予測フレーズ長：フレーズの平均的な時間長で、 T （単位は ms ）とする。（本実施例では $T = 2000 \text{ ms}$ とした）また、図6中の $f(t)$ と $g(t, w)$ は有声区間の伸張倍率を定める関数であり、以下の特性を有するものである。

【0047】 $f(t)$ ：話速変換に伴う時間伸張を吸収するために用いる倍率関数であって、予測フレーズ長内の有声区間の出現時刻 t ($0 \leq t \leq T$) に対して倍率を定める単調減少関数である。

【0048】 $t = 0$ におけるあらかじめ定めた倍率を r_s 、 $t = T$ におけるあらかじめ定めた倍率を r_e ($r_s \geq r_e$) とすると、 $f(t)$ は $r_s \geq f(t) \geq r_e$ 、 $0 \leq t \leq T$ を満たす。

【0049】 $g(t, w)$ ：フレーズの開始部分から一定時間 T_0 内は、 $g(t, w)$ を適用し、有声区間の伸張倍率を、その出現時刻 t と区間長 w の双方を考慮して、 $f(t)$ により定まる規準倍率より高い倍率で伸張することにより、各音韻長の大小関係を考慮しつつ、フレーズ開始点での「ゆっくり感」を強調するための倍率関数である。

【0050】次に、図5の処理手順を説明する。なお、 ST はステップを意味する。

【0051】($ST0$) まず、 $f(t)$ の最高倍率 r_s と最低倍率 r_e を設定する。

【0052】($ST0-1$) 次に、フレーム番号 i を0にセットする。

【0053】($ST0-2$) 続いて、上記 i を $i+1$ とインクリメントする。

【0054】($ST1$) そして、音声入力回路1が取り込んだ入力音声を、フレームと呼ばれる一定長の部分に分割し、その結果を入力バッファ回路4に格納する処理を行う。

【0055】本実施例ではフレーム幅 6.6 ms の Hamm ing (ハミング) 窓を 3.3 ms ずつずらしながら切り出して格納する。

【0056】($ST2$) 入力音声信号を各フレーム毎に、自己相関法や、零クロス法などの方法で処理して有声、無声、無音の判定を行う。人が発声する有声および無声以外の入力音（例えば、低レベルの雑音や背景音等）は原則として無音として識別処理する。

【0057】($ST3$) i 番目のフレームについての有声、無声、無音の判定結果（今回の判定結果）と、 $i-1$ 番目のフレームについて有声、無声、無音の判定結果（前回の判定結果）とが同じであるか否かを判別する。両者の判定結果が同じであれば ($ST0-2$) に戻り、同じでないならば次の ($ST4$) に移る。

【0058】($ST4$) $i-1$ フレームまでの、同じ種類（有声、無声或いは無音）の区間と判定されている音声区間を入力バッファ回路4から処理バッファ回路5に転送して格納する。

【0059】($ST5$) 処理バッファ回路5に格納されている音声区間が、無音か無声か有声か否かを判定する。無音区間の場合は ($ST6$) へ進み、無声区間の場合は ($ST11$) へ移り、有声区間の場合は ($ST9$) へ移る。

【0060】($ST6$) 当該無音区間がポーズ区間か否かを判断する。ポーズ区間の場合は ($ST6-1$) へ移り、ポーズ区間でない場合は ($ST8$) へ飛ぶ。但し、図4のリアルタイム話速変換装置の起動時はポーズ区間であったと判断し、必ず ($ST6-1$) へ進む。

【0061】($ST6-1$) ポーズ区間以降に出現する有声区間の番号を表す変数 k に初期値としての1を代入する。

【0062】($ST7$) ポーズの区間長を調べ、その区間長によって適宜、予め設定されているアルゴリズム（池沢龍ほか「話速変換に伴う時間伸張を吸収するための一方法」1992年音声研究会P. 49-P. 56）によって聴感上違和感ない程度に短縮する。

【0063】本実施例では、 862 ms を越える区間長を有する無音区間を一律にこの 862 ms の値まで短縮することとし（池沢龍ほか平成4年日本音響学会春季研究発表会「話速変換に伴う時間伸張を吸収するための一手法」2-6-2 P. 331-P. 332 (1992-3)）、無音区間 862 ms を経過した時点で更に無音区間が続く場合は、それ以降の無音データを廃棄して次のフレーズの開始点を待つものとする。

【0064】($ST8$) 処理バッファ回路5内にある処理済の無音区間の信号をファイル回路6に転送させて格納させた後、処理バッファ回路5をクリアする。次に ($ST12$) へ移る。

【0065】($ST12$) 音声信号の最後まで処理したか否かを判定する。肯定判定の場合は本 ($ST9$) 処理（ルーチン）を終了し、否定判定の場合は ($ST0-2$) へ戻る。

【0066】($ST9$) ($ST5$) で有声区間と判定された区間に対して、後述の図6に示す有声区間処理を行う。 $f(t)$ の適用開始時刻を示す変数として V_{st} を導入する。また、フレーズ内の第 k 有声区間の開始時刻を t_i 、区間長を w_i と記す。

【0067】($ST9-1$) 上述の変数 k を $k+1$ とイ

ンクリントする。

【0068】(ST10) 処理バッファ回路5内にある話速変換済みの音声データをファイル回路6のメモリに格納するとともに、処理バッファ回路5をクリアする。その後、上述の(ST12)へ移る。

【0069】(ST11) (ST5)において処理対象となる区間が無声と判断されれば、この無声区間の音声信号を処理バッファ回路5からファイル回路6に転送して格納した後、処理バッファ回路5をクリアする。その後、上述の(ST12)へ移る。

【0070】次に、図6のST9有声区間処理ルーチンの詳細を説明する。

【0071】(ST14) まず、有声区間のピッチ抽出を行う。

【0072】(ST15) 次に、変数kがk=1か否かを判定する。k=1の場合、即ちポーズ区間以降に出現する最初の有声区間の場合は(ST15-1)へ移り、そうでない場合は(ST15-2)へ移る。

【0073】(ST15-1) 変数V_{st}に時刻t_iを代入する。次に(ST16)へ移る。

【0074】(ST15-2) 変数kが3以下か否か、即ちkが2または3であるか否かを判定する。kが2または3の場合は(ST16)へ移り、kが4以上の場合は(ST17)へ飛ぶ。

【0075】(ST16) 第k有声区間の最大ピッチ周波数をP_kと定義する。k=1, 2, 3の場合にはP_kの値を保存する。

【0076】(ST16-1) 変数kがk=3か否かを判定する。k=3の場合は次の(ST16-2)へ移り、そうでない場合、即ちk=1, 2の場合は(ST17)へ飛ぶ。

【0077】(ST16-2) 3つの有声区間P₁, P₂, P₃のうちの最大値をPinch_{max}

*

$$f(t) = r_e + 0.5(r_s - r_e) \{ \cos \pi (t - V_{st}) / T + 1.0 \} \quad (1)$$

但し、 $V_{st} \leq t \leq V_{st} + T$

例えば、k番目の有声区間に対する伸張倍率はf(t_k)となる。その後、図5のメインルーチンに戻る。

【0084】(ST18-1) k番目の有声区間長w_kと、予め設定されている区間長w₀が、w_k ≤ w₀であるか否かを判定する。

【0085】この条件を満たせば(ST18-2)へ、そうでなければ(ST18-4)へ移る。

【0086】w₀は1つの有声区間内での分割区間長で、区間長がw₀以上の有声区間は、全区間に渡って一様な倍率で伸張するのではなく、この分割区間単位で倍率を変更(順次減少させる)することにより、話速変換による、1フレーズ内での音韻長の時間的大小関係を保ち、同時に「まのび感」を抑制することができる。

【0087】本実施例では、w₀ = 150msとした。

*とする。次に(ST17)へ移る。

【0078】(ST17) t_iが、区間[V_{st}, V_{st} + T]に含まれているか否かを判定する。含まれていれば(ST17-1)へ移り、そうでなければ(ST21)へ移る。(本実施例では前述のようにT=2000msとした。)

(ST17-1) V_{st} > t_iであるかを判定する。

【0079】V_{st} > t_iであれば(ST19)へ移り、そうでなければ(ST18)に移る。V_{st} > t_iのときは、発声の終了点間近で意味的重要度が低い場合が多いため、本実施例では特にg(t, w)を適用せず、(ST17-1)から直接(ST19)に移ることとした。

【0080】(ST18) フレーズの開始部において変換により生じる聴感上の「ゆっくり感」を効果的にするために必要な時間長をT₀とする。T₀は実験結果から(今井篤ほか平成5年日本音響学会秋季研究発表会「話速変換に伴う時間伸張のリアルタイム吸収法」1-9-10P. 361-P. 362(1993-10)) T₀の1/4程度が望ましく、本実施例ではT₀ = 450msとした。

【0081】本処理ブロックでは、第k有声区間の終了時刻t_k + w_kが区間[V_{st}, V_{st} + T₀]に含まれているか否かを判定する。含まれていれば次の(ST18-1)へ移り、そうでなければ(ST19)へ移る。

【0082】(ST19) 予め設定した倍率関数f

(t)を適用して有声区間を伸張する。このf(t)は単調減少関数であり、本実施例では以下の式(1)のような余弦関数を用いて、倍率をr_eからr_sまで変化させた。

【0083】

【数1】

【0088】(ST18-2) 変数f_{lg}に1を代入し、(ST20)へ移る。

【0089】(ST20)へ移る。

【0090】(ST18-4) 変数f_{lg}に0を代入し、(ST20)へ移る。

【0091】(ST20) 予め設定した関数g(t, w)を適用して有声区間を伸張する。

【0092】g(t, w)の適用開始時刻を示す関数としてB_{st}を導入する。但し本実施例ではB_{st} = t_iとする。

【0093】g(t, w)は単調減少関数であり、区間[B_{st}, B_{st} + T₀]において常にg(t, w) ≥ f(t)が成り立つことを条件とする。本実施例は

【0094】

【数 2】

$$t_k + w_k \leq B_st + T_0 < t_{k+1} + w_{k+1} \quad (2)$$

であるとき、第 k 有声区間を $g(t_k, w_k)$ で、第 k+1 有声区間を $f(t_{k+1})$ で伸張する場合に、第 k+1 有声区間の伸張倍率が第 k 有声区間の伸張倍率より大きくなるためのものである。この条件を満たしていれば、 $g(t, w)$ の適用区間 $[B_st, B_st + T_0]$ において、減少倍率の逆転は起こらない。

【0095】本実施例で用いた倍率関数 $g(t, w)$ を*

$$g(t_k, w_k) = \begin{cases} -((r_k^2 - f(T_0))(t_k + w_k - B_st)/T_0) + r_k^2 & \text{但し、} flg=1 \\ -((r_k^2 - f(T_0))(t_k + w_k - B_st)/T_0) + r_k^2 & \text{但し、} flg=0 \end{cases}$$

(3)

但し、 $0 < T_0 < T$,

$$B_st < t < B_st + T_0$$

【0097】(ST20-1) 次に、 $g(t, w)$ に従って伸張した有声区間を処理バッファ回路 5 に格納する。

【0098】(ST20-2) 次に、 $flg=1$ であるか否かを判定する。 $flg=1$ のときは (ST9) の有声区間処理ルーチンを終了する。 $flg=0$ のときは、分割された有声区間が更に続くということであり、(ST20-3) に移る。

【0099】(ST20-3) t_k の値を $t_k + w_k$ に*

$$P_k > Pitch_max \times Th2$$

本実施例では、 $Th2 = 0.7$ とした。

【0103】(ST22) 変数 V_st に時刻 t_k を代入する。

【0104】(ST22-1) 変数 r に $(r, -Th3)$ を代入する。

【0105】これによって、 $f(t)$ は $(r, -Th3)$ から r まで倍率を変化させる。本実施例では、 $Th3 = 0.1$ に設定した。その後、上記の (ST17) へ戻る。

【0106】(ST23) 有声区間を伸張倍率 r で伸張する。つまり、話速を最も速い状態のままにする。その後、(ST9) の有声区間処理ルーチンを終了し、図 5 のメインルーチンに戻る。

【0107】

【発明の効果】以上説明したように、本発明によれば、入力音声の無音区間、無声区間、有声区間を分離し、有声区間を一定の規則に従って伸張することによって、所望の聴感的効果を安定に得ることが可能な話速変換方法において、入力音声の話速を規準となる所望の倍率によって一定時間内に「ゆっくり」の状態から「速い」状態に変換する際に、従来方法のように有声区間の出現時刻

* 以下の次式 (3) に示す。一次関数を採用し、時刻 $B_st + T_0$ の時に $f(t)$ の倍率に一致するようにした。第 k 有声区間の出現時刻 t_k と、その区間長 w_k を用いて、

【0096】

【数 3】

20※ 更新する。

【0100】(ST20-4) 次に、 w_k の値を $w_k - w$ に更新し、(ST18-1) へ戻る。

【0101】(ST21) 処理対象となっている有声区間の最大ピッチ周波数 P_k が、以下の式 (4) の条件を満たす場合は (ST22) へ、満たさない場合は (ST23) へ移る。

【0102】

【数 4】

(4)

30 によって倍率を一意的に定めるのではなく、有声区間の区間長とその出現時刻の双方を加味してその伸張倍率の値を変化させる倍率変数を適用するようにしているの

で、規準となる倍率変数の変化に整合した聴感的効果が得られ、且つ、音韻長の時間的な大小関係を乱すことなく、所望の話速変化の効果を得ることができる。

【0108】従って、本発明によれば、話速を「ゆっくり」から「速く」に変化させる場合において、受聴者の希望にあった話速変化を安定、且つ自然に実現することができる。

40 【図面の簡単な説明】

【図 1】従来法における倍率関数を示すグラフである。

【図 2】従来法を適用した場合の 1 フレーズ内の有声区間長の時間軸上の分布を示すタイミング図である。

【図 3】本発明の一実施例の倍率関数を示すグラフである。

【図 4】本発明によるリアルタイム話速変換方法の一実施例を適用したリアルタイム話速変換装置の回路構成例を示すブロック図である。

50 【図 5】図 4 に示すリアルタイム話速変換装置の動作例を示すメインフローチャートである。

【図6】図5に示す有声区間処理ルーチンの詳細を示すフローチャートである。

【図7】図4に示すリアルタイム話速変換装置に $f(t)$ 、 $g(t, w)$ の関数を適用した場合の動作例を示すタイミング図である。

【符号の説明】

- 1 音声入力回路
- 2 CPU回路
- 3 PROM回路
- 4 入力バッファ回路
- 5 処理バッファ回路
- 6 ファイル回路
- 7 音声出力回路
- 8 バス

$f(t)$ 話速変換に伴う時間伸張を吸収するために用いる倍率関数

* $g(t, w)$ 区間長 w とその出現時刻 t に応じて、 $f(t)$ により定まる規準倍率より高い倍率で伸張するための倍率関数

r_s あらかじめ定めた最高倍率

r_e あらかじめ定めた最低倍率

T 予測フレーズ長（フレーズの平均的な時間長）

Ph_st フレーズ（ポーズと次のポーズに挟まれる区間）の開始点

V_st $f(t)$ の適用開始時刻を示す変数

10 P_k 第 k 有声区間の最大ピッチ周波数

$pitch_max$ 最初の3つの有声区間 P_1 、 P_2 、 P_3 のうち最大値

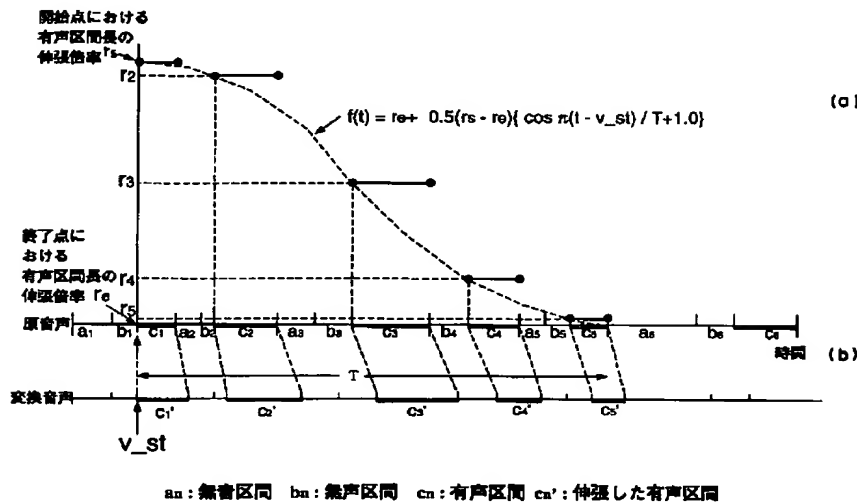
w_k k 番目の有声区間長

i フレーム番号

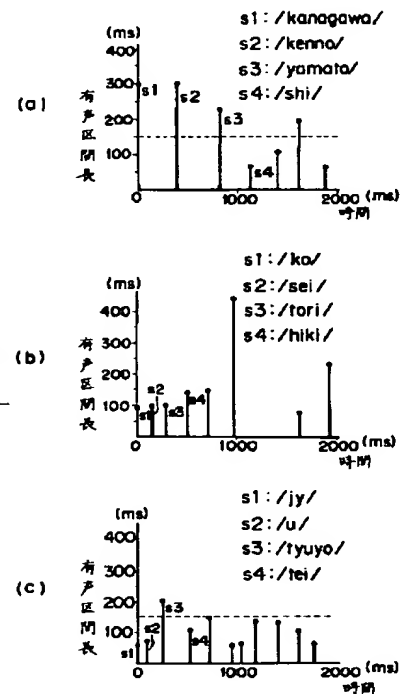
k 有声区間番号

* t_k 第 k 有声区間の開始時刻

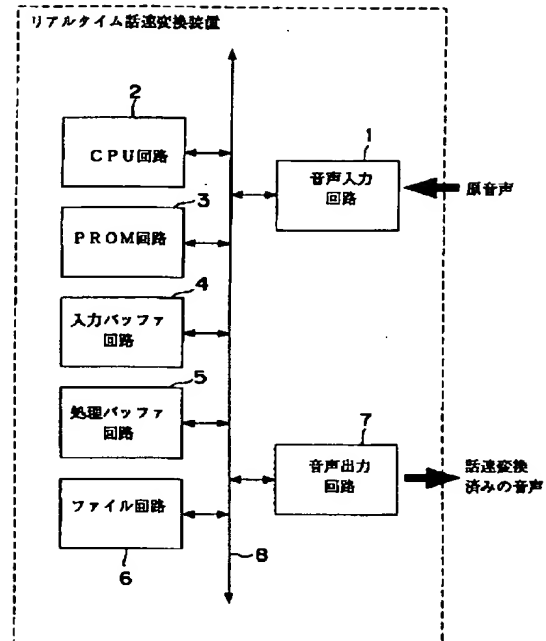
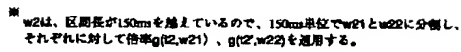
【図1】



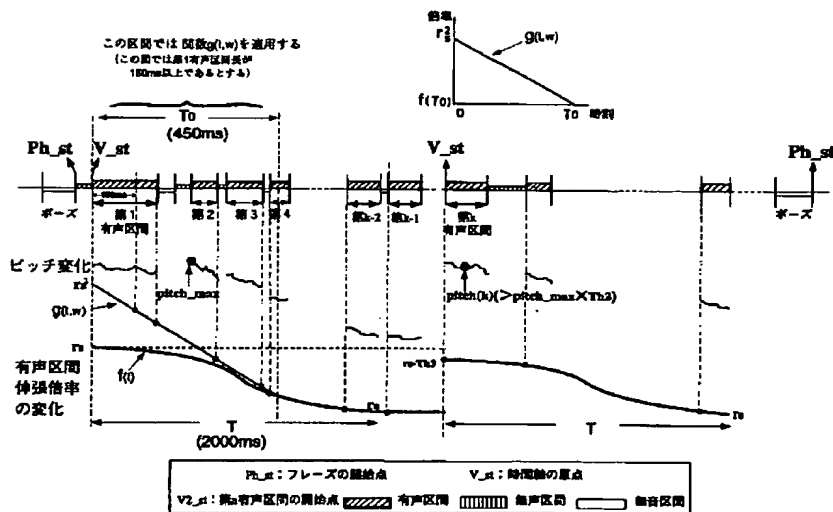
【図2】



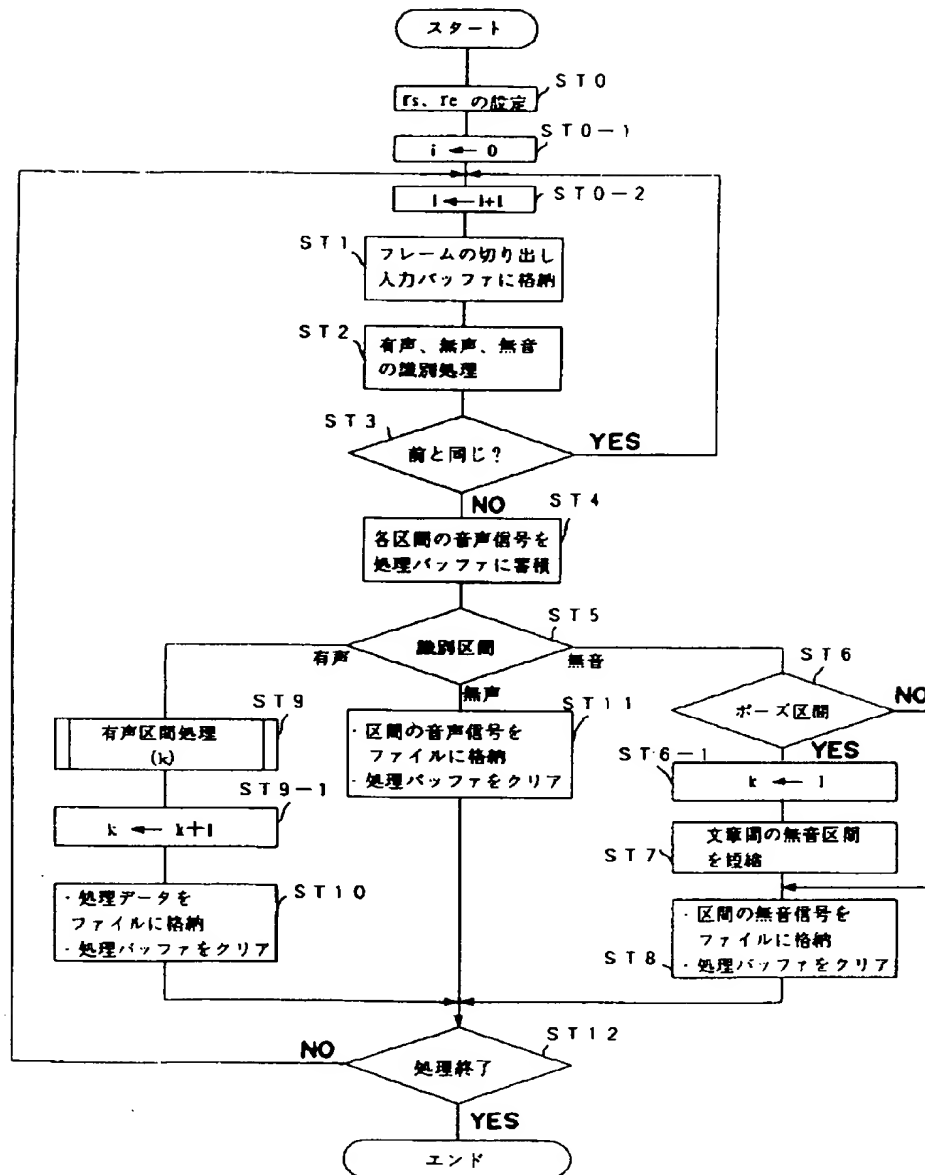
【図 4】



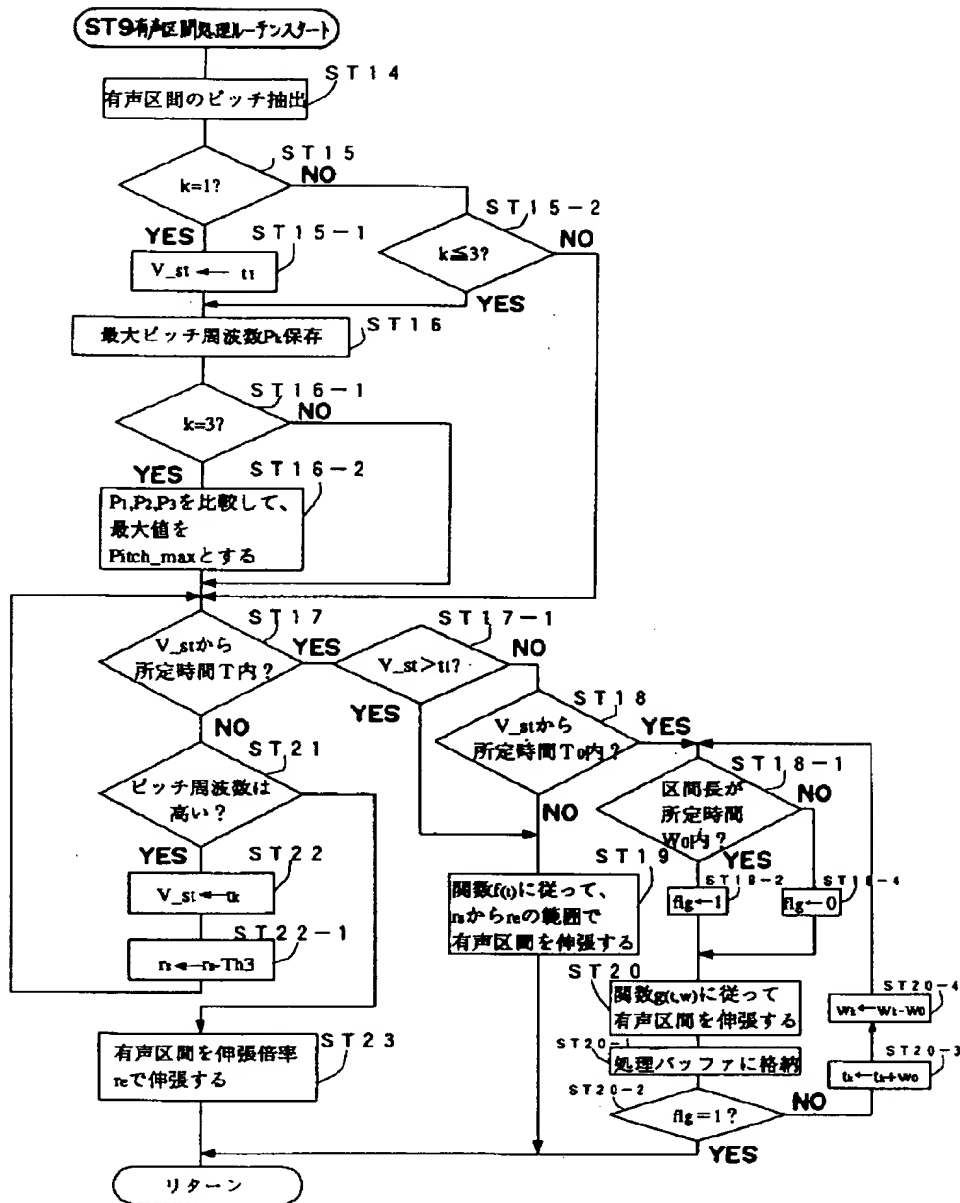
【図 7】



【図5】



【図6】



フロントページの続き

(72)発明者 清山 信正
 東京都世田谷区砧一丁目10番11号 日本放
 送協会 放送技術研究所内

(72)発明者 宮坂 栄一
 東京都世田谷区砧一丁目10番11号 日本放
 送協会 放送技術研究所内

【公報種別】特許法第 17 条の 2 の規定による補正の掲載
 【部門区分】第 6 部門第 2 区分
 【発行日】平成 11 年（1999）8 月 6 日

【公開番号】特開平 7-281691
 【公開日】平成 7 年（1995）10 月 27 日
 【年通号数】公開特許公報 7-2817
 【出願番号】特願平 6-67251
 【国際特許分類第 6 版】
 G10L 3/00
 【F I】
 G10L 3/00 H

【手続補正書】
 【提出日】平成 10 年 8 月 25 日
 【手続補正 1】
 【補正対象書類名】明細書
 【補正対象項目名】発明の名称
 【補正方法】変更
 【補正内容】
 【発明の名称】 話速変換装置
 【手続補正 2】
 【補正対象書類名】明細書
 【補正対象項目名】特許請求の範囲
 【補正方法】変更
 【補正内容】
 【特許請求の範囲】

【請求項 1】 入力音声の無音区間、無声区間、有声区間を分離し、このうち有声区間を伸張することによって発声の速さ（話速）を声の高さを保ったまま遅くする変換を行う際に、各有声区間の時間長を逐次検出し、各々の有声区間の時間長に一樣な値の、あるいは経過時間とともに滑らかに変化する規準倍率を乗ずることにより、その倍率に対応した聴感的な効果を得る話速変換装置であって、一息で発声される区間（フレーズ）の一定時間内において、話速を「ゆっくり」の状態から「速い」状態に変化させる場合に、望まれる聴感上の話速変換効果において、1 モーラ（1 音）毎のテンポが自然に変化するように、各有声区間の時間長と、その有声区間の出現時刻の双方を考慮した一定の規則に基づいて、前記規準倍率に比べてより高い倍率を決定する決定手段を有することを特徴とする話速変換装置。

【請求項 2】 前記経過時間とともに滑らかに変化する規準倍率として、一息で発声する区間を単位にしてこの区間の開始点ではゆっくりとした話速を設定し、その終了点に向かって徐々に話速を速めることを特徴とする倍率関数を適用する場合に、上記区間の開始時刻から時間 450ms 以内に出現する 150ms 以下の短い有声区間に対しては、前記一定の規則に基づいて前記規準倍率

に比べてより高い伸張倍率を供する倍率関数に沿って、その有声区間の時間長とその有声区間の出現時刻の双方を加味した伸張倍率を乗じ、また前記 150ms を越える時間長の有声区間に対しては、該 150ms 単位で分割し、各々の終了点に対応する時刻での伸張倍率を乗じ、さらに出現時刻が 450ms を越える有声区間に対しては前記規準倍率を乗ずる演算手段を有することを特徴とする請求項 1 に記載の話速変換装置。

【請求項 3】 前記一定の規則は、変換開始直後に出現する有声区間の伸張倍率が最も高く、時間経過と共に倍率は減少するが、相次いで現れる有声区間については、その時間長が短いものほど倍率の減少率は小さくするという規則であることを特徴とする請求項 1 に記載の話速変換装置。

【請求項 4】 前記一定の規則は、任意の隣合う k 番目、および $k+1$ 番目の有声区間の時間長をそれぞれ w_k 、 w_{k+1} とし、話速変換後の時間長をそれぞれ w'_k 、 w'_{k+1} とすると、 w_k 、 w_{k+1} が近い値で且つ $w_k \leq w_{k+1}$ である場合のみ、 $w'_k > w'_{k+1}$ となることを許容することを特徴とする請求項 1 に記載の話速変換装置。

【請求項 5】 前記一定時間は、ポーズ直後またはピッチの変化がある一定の値を越えた時刻から、フレーズの終了時刻までの間で任意に定める時間であり、所望の効果によってその時間の値を任意に決定できることを特徴とする請求項 1 に記載の話速変換装置。

【請求項 6】 前記一定の規則は、当該有声区間の時間長が、ある一定の値よりも大きい場合には、所定の時間長単位で分割し、分割されたそれぞれの有声区間の時間長とその分割時刻の双方を考慮してそれぞれの倍率を決定するという規則であることを特徴とする請求項 1 に記載の話速変換装置。

【請求項 7】 前記所定の時間長は、実用的な値の規準倍率によって有声区間を伸張したときに、聴感的な効果が感じとれなくなる最大の有声区間の時間長であることを特徴とする請求項 6 に記載の話速変換装置。

【手続補正3】

【補正対象書類名】明細書

【補正対象項目名】0001

【補正方法】変更

【補正内容】

【0001】

【産業上の利用分野】本発明は、話速変換装置に関し、特に聴覚障害者や高齢者等の音声補聴装置や、一般的な語学学習装置、ラジオ、テープレコーダー、電話などにおいて、話速変換による補助的聴取を行う際の聞き取り易さの向上、テレビジョン、ビデオテープレコーダー、ビデオディスクプレーヤーなどの音声出力を話速変換した際に生ずる映像と音声のズレを効果的に吸収するリアルタイム式の話速変換装置に関する。

【手続補正4】

【補正対象書類名】明細書

【補正対象項目名】0016

【補正方法】変更

【補正内容】

【0016】本発明は、上述の問題点に鑑みてなされたもので、その目的は任意に定めた一定時間内において話速を「ゆっくり」の状態から「速い」状態に変化させる場合に、各音韻間の大小関係を保存し、各有声区間の変換音声を設定された話速に相応の聴感的効果を保ちつつ、安定した話速変換効果を得られるようにし、これにより聴覚障害者や高齢者に施す話速変換による補聴効果の向上や、定まった時間枠で出来るだけ自然なゆっくりとした音声を提供することが可能な話速変換装置を提供することにある。

【手続補正5】

【補正対象書類名】明細書

【補正対象項目名】0017

【補正方法】変更

【補正内容】

【0017】

【課題を解決するための手段】上記目的を達成するために、本発明は、入力音声の無音区間、無声区間、有声区間を分離し、このうち有声区間を伸張することによって発声の速さ（話速）を声の高さを保ったまま遅くする変換を行う際に、各有声区間の時間長を逐次検出し、各々

の有声区間の時間長に一樣な値の、あるいは経過時間とともに滑らかに変化する規準倍率を乗ずることにより、その倍率に対応した聴感的な効果を得る話速変換装置であって、一息で発声される区間（フレーズ）の一定時間内において、話速を「ゆっくり」の状態から「速い」状態に変化させる場合に、望まれる聴感上の話速変換効果において、1モーラ（1音）毎のテンポが自然に変化するように、各有声区間の時間長と、その有声区間の出現時刻の双方を考慮した一定の規則に基づいて、前記規準倍率に比べてより高い倍率を決定する決定手段を有することを特徴とする。

【手続補正6】

【補正対象書類名】明細書

【補正対象項目名】0018

【補正方法】変更

【補正内容】

【0018】また、好ましくは、本発明は、前記経過時間とともに滑らかに変化する規準倍率として、一息で発声する区間を単位にして、この区間の開始点ではゆっくりとした話速を設定し、その終了点に向かって徐々に話速を速めることを特徴とする倍率関数を適用する場合に、上記区間の開始時刻から一定時間内、好ましくは時間450ms以内に出現する前記所定の長さに当る区間長、好ましくは150ms程度に満たない区間長を有する有声区間に対しては、前記一定の規則に基づいて前記規準倍率に比べてより高い伸張倍率を供する倍率関数に沿って、その有声区間の時間長とその有声区間の出現時刻の双方を加味した伸張倍率を乗じ、また前記150msを越える時間長の有声区間に対しては、該150ms単位で分割し、各々の終了点に対応する時刻での伸張倍率を乗じ、さらに出現時刻が450msを越える有声区間に対しては前記規準倍率を乗ずる演算手段を有することを特徴とすることができる。

【手続補正7】

【補正対象書類名】明細書

【補正対象項目名】図4

【補正方法】変更

【補正内容】

【図4】本発明の一実施例のリアルタイム話速変換装置の回路構成例を示すブロック図である。